

CRC Tissue Core Management System (TCMS): Integration of Basic Science and Clinical Data for Translational Research

Andrew W. Greaves, B.S.; Philip R.O. Payne, M.A.; Laura Rassenti, Ph.D.; Thomas J. Kipps, M.D. Ph.D.

UCSD Rebecca and John Moores Cancer Center, La Jolla, California

Abstract: The Chronic Lymphocytic Leukemia (CLL) Research Consortium (CRC) consists of 9 geographically distributed sites conducting a program of research including both basic science and clinical components. The CRC TCMS was designed to capture and integrate basic science and clinical data sets. The system utilizes multiple data modeling methodologies and web-application platforms, and was designed with the high level objectives of providing an extensible, generalizable model for integrating data as required to conduct translational research.

The CLL Research Consortium (CRC) is a multi-institutional research program funded by the National Cancer Institute (PO1 CA81534) and headquartered at the UCSD Rebecca and John Moores Cancer Center in La Jolla, California. The consortium consists of 8 additional sites including The Burnham Institute (La Jolla, CA), The Dana Farber Cancer Institute (Boston, MA), Johns Hopkins University (Baltimore, MD), Long Island Jewish Medical Center (New Hyde Park, NY), M.D. Anderson Cancer Center (Houston, TX), The Ohio State University Cancer Center (Columbus, OH), Thomas Jefferson University (Philadelphia, PA), and the Mayo Clinic (Rochester, MN). The consortium conducts an integrated program of basic science and clinical research with the objective of determining curative strategies for Chronic Lymphocytic Leukemia

In order to facilitate the collaborative collection, distribution, and analysis of data and tissue samples by the consortiums multiple sites, an information management solution that addressed both the administrative and research oriented data management tasks inherent to such activities was required. The resulting system, TCMS, has been designed to not only capture multiple, disparate data sets, but also, to allow CRC participants to utilize such information via a single conceptual model. Furthermore, the TCMS web application closely integrates with the CRC's clinical trials management system (CTMS), providing a means of enabling correlative basic science endeavors during the course of clinical protocols. As of the time of submission of this abstract, in excess of 2000 patients and over

68,000 tissue samples have been accessioned and are presently managed via TCMS.

The primary features of TCMS are web-based basic-science data capture, tissue sample management, and integrated data mining tools that span both the TCMS and CTMS data repositories. These features are described in greater detail in the following sections:

Web-based Basic Science Data Capture:

Utilizing a hybrid data model incorporating an entity-attribute-value (EAV) data repository and multiple relational data-marts, the real-time definition and/or re-definition of the TCMS data dictionary is made possible. This is of the utmost importance in allowing basic scientists within the CRC to define and submit novel data sets via the web-based TCMS interface without development staff intervention.

Tissue Sample Management:

All data related to the submission of tissue samples to the centralized CRC tissue core is captured via the TCMS web interface. Internal correlative data (ex. flow cytometry data) is captured via automated data management systems and stored in the TCMS data repository. Requests for samples from the CRC tissue core are submitted via the TCMS web-site, and outgoing samples are physically tracked using a bar-coding system.

Integrated Data Mining Tools:

Queries spanning both the TCMS data set and the related CTMS data set are made possible through the use of an integrated EAV data warehouse and a combination of both thick-client and web-based query tools. Sub-systems within the TCMS provide for time-normalization of variable data sets as well as population based statistics, enabling such data mining tasks.

Future directions for TCMS include the development of ad-hoc query tools utilizing XML/XSLT-based data export and translation in order to provide real-time data-set availability for the purposes of statistical analysis using any number of commercial software packages. In addition, the developers intend to further integrate TCMS with the CTMS web-application patient tracking features.